

# On the Estimation of Survival of HIV/AIDS Patients on Antiretroviral Therapy Using NPMLE Method: An Application to Interval Censored Data

Gurprit Grover<sup>1</sup>, Rabindranath Das<sup>2</sup>, Prafulla Kumar Swain<sup>1,\*</sup>, Barnali Deka<sup>1</sup>

<sup>1</sup>Department of Statistics, University of Delhi, Delhi, India

<sup>2</sup>Department of Statistics, University of Burdwan, West Bengal, India

**Abstract** The main objective of this paper is to estimate the survival of HIV/AIDS patients who are undergoing Antiretroviral Therapy in an ART centre, Delhi. Non Parametric Maximum Likelihood Estimation NPMLE (E-M) for interval censoring and KM survival plot for left, right and mid-point imputation have been used to estimate the survival of these patients. It has been observed that the mid-point imputed survival plot has a very similar and consistent pattern as obtained by NPMLE (E-M) method. Considering these mid-point imputed value as right censored data, Cox PH model and Accelerated Failure time Model (AFTM) have been applied to identify the effects of prognostic factors like age, sex, mode of transmission, baseline CD4 cell count, hemoglobin, baseline weight and smoking habits on survival of the patients. The Akaike Information Criterion (AIC) has been employed to compare the efficiency of these models and Cox-Snell residual is used to test the proportionality assumption.

**Keywords** AIDS, ART, CD4 Cell Count, Non Parametric Maximum Likelihood Estimation, Cox-PH Model, Accelerated Failure Time Model

## 1. Introduction

The widespread use of Antiretroviral Therapy (ART), prognosis of HIV infected patients has significantly improved and mortality due to AIDS related causes has substantially reduced [1, 2, 3]. With the advent of free ART programme in 2004 in India, a dramatic decline of HIV incidence has been observed during the last decade [4]. Thus it is imperative to know the survival of HIV/AIDS patients on ART, so that appropriate strategies can be devised to prevent new infection.

However in HIV/AIDS studies, where time to event data are collected by assessing patients in periodic follow up visits. In such cases, the event cannot be observed exactly (i.e onset of HIV infection, incubation period of AIDS after HIV infection, death time); however it is known to happen within some interval, thus such observed events are interval censored. In HIV dynamics every patient is supposed to visit ART centre after four weeks, but actual visit time vary from patient to patient and also time between visits vary. The patients may visit ART centre at time that is convenient to them rather than scheduled time. Therefore the failure event

(i.e death) is considered as interval censored survival data. Survival time  $T_i$  (say) of a patient lies in the interval  $(L_i, R_i)$  that is the last available visit date and end of the study.

Several approaches have been proposed for the estimation of survival function for the interval censored data; Turnbull [5], De Gruttola and Lagakos [6] and Sun [7, 8]. More recently Kim [9], Grover and Shakeri [10], Grover and Banerjee [11] estimated survival of HIV-1 infected children for doubly and interval censored data.

In this paper we have estimated the survival function of HIV/AIDS patients on ART by using Non Parametric Maximum Likelihood Estimation (NPMLE) method for interval censored data. The NPMLE is computed by using the E-M algorithm of Turnbull (1976) with the polishing algorithm of Gentleman and Geyer [12]. We have also estimated survival function by using imputation technique. Due to lack of appropriate estimation procedure and non availability of statistical software for interval censored data, we have adopted conventional imputation approach to convert interval censored data to right censored data, for which the standard techniques are available. Further, we have compared the estimates obtained by NPMLE and imputation method.

Kaplan-Meier plot for left, right and mid-point imputation has been used to estimate the survival of patients. It has been observed that the mid-point imputed survival plot has a very similar and consistent pattern as obtained by NPMLE

\* Corresponding author:

prafula86@gmail.co (Prafulla Kumar Swain)

Published online at <http://journal.sapub.org/ajms>

Copyright © 2013 Scientific & Academic Publishing. All Rights Reserved

(E-M) method. Considering these mid-point imputed values as right censored data at each interval, Cox Proportional Hazard model and Accelerated Failure Time Model (AFTM) has been used to study the effect of prognostic factors like age, sex, mode of transmission, baseline CD4 cell count, hemoglobin, baseline weight and smoking habits on the survival of patients.

The semi parametric Cox PH model is the most common approach in survival analysis and has been used to evaluate the covariate effects on hazard function of failure time data. Interestingly, the result from a PH model is difficult to interpret in terms of survivorship. However, parametric AFTM is being treated as best attractive alternative to Cox PH model. AFT model provides concise and more intuitively interpretable results of survival data[13]. Since no one has used AFTM in HIV population studies in India earlier, we have modeled HIV/AIDS population on ART by using AFTM and compared the results with Cox PH model.

The Akaike Information Criterion (AIC) has been employed to compare the efficiency of models. Schoenfeld residual and Cox-Snell residual method has also been used to test proportionality assumption. The software packages *survival and interval* in R and STATA(version 11.1) have been used to perform the statistical analyses.

## 2. Methods Used

### 2.1. NPMLE for Interval Censored Data

Suppose that  $n$  number of HIV/AIDS patients are under ART, Let  $T_i$  be the survival time (i.e occurrence of death) of  $i^{\text{th}}$  patient with survival function  $S(t)$ . Let  $(L_i, R_i)$  be the interval in which  $T_i$  is being observed, such that  $L_i < T_i < R_i$ , and if the event does not occur till the end of study then the patient is said to be right censored. In this case we assume that  $T_i$  can occur in the interval  $(L_i, \infty)$ , where  $L_i$  is the time period from the beginning of the study until the last visit. The likelihood function for the set of observed intervals  $\{(L_i, R_i), i = 1, 2, \dots, n\}$  is then given by

$$L = \prod_{i=1}^n \{S(L_i) - S(R_i)\} \quad (1)$$

Turnbull's EM algorithm approach, based on iterative procedure has been used to estimate the survival function  $S(t)$  corresponding to interval censored data. From the data  $\{(L_i, R_i), i = 1, 2, \dots, n\}$ , a set of non overlapping intervals  $\{(t_1, t_2), \dots, (t_{m-1}, t_m)\}$  is generated over which the survival curve  $S(t)$  is estimated. Let  $0 = t_0 < t_1 < \dots < t_m = \infty$  be the ordered observed survival time where the NPMLE may change. Then for  $i^{\text{th}}$  patient  $\{T_i \in (L_i, R_i), i = 1, 2, \dots, n\}$ , define an indicator variable

$$I_{ij} = \begin{cases} 1 & \text{if } (t_{j-1}, t_j) \in (L_i, R_i) \\ 0 & \text{otherwise} \end{cases}$$

and  $P(t_j) = P[t_{(j-1)} < T_i < t_j], j = 1, 2, \dots, m$ . Then the likelihood function  $L(P)$  under constraint  $\sum_{j=1}^m P(t_j) = 1$ , is given by

$$L(P) = \prod_{i=1}^n \sum_{j=1}^m I_{ij} P(t_j) \quad (2)$$

Where  $P_i(t_j) = P[t_{(j-1)} < T_i < t_j / T_i \in (L_i, R_i)]$ , Now the NPMLE of survival function can be estimated by

maximizing  $L(P)$  with respect to  $P$ , say  $\hat{P}$ , by using EM algorithm[14]. Starting with initial estimate  $\hat{P}$ ,

E-step, for each  $i$

$$\hat{P}_i(t_j) = \frac{P_i(t_j) I\{t_j \in (L_i, R_i)\}}{\sum_{tk \in (L_i, R_i)} \hat{P}(tk)}$$

M-step: update  $\hat{P}$  for each  $j$

$$\hat{P}(t_j) = \frac{\sum_{i=1}^n \hat{P}_i(t_j)}{n}$$

Iterate the procedure until convergence.

### 2.2. Imputation Approach and K-M Survival Plot

For simplicity, imputation techniques have been used to handle this interval censored data. After imputation the interval censored data will be treated as complete or right censored data, then standard statistical methods can be performed on the imputed data set (Hsu et al.,[15]).[16, 17] have used mid-point imputation for interval censored AIDS infection time data. Since the survival time  $T_i$  is known to lie in the interval  $\{(L_i, R_i), i = 1, 2, \dots, n\}$  for  $i^{\text{th}}$  patient. we use the midpoint ( $m_i$ ) of the observed interval as the hypothetical failure time i.e  $m_i = \frac{(L_i + R_i)}{2}$ . And also an alternative to midpoint imputation is to take  $T_i$  as  $L_i$  the left end point and  $R_i$  the right end point imputation have been used.

Kaplan-Meier survival estimate have been plotted for this imputed data, by using the formula

$$\widehat{S}(t) = \prod_{j: t_j \leq t} (1 - \frac{d_j}{n_j}) \quad (3)$$

where the  $t_j$ 's denote the distinct imputed exact failure times,  $d_j$ 's and  $n_j$ 's are the death and number at risk at each of the  $t_j$ 's respectively.

### 2.3. Cox-PH Model

The Cox proportional hazard function[18] for the survival time  $T$  and the explanatory variables  $X$  is given by

$$h(t/X) = h_0(t) \exp(\beta' X) \quad (4)$$

where  $h_0(t)$  is the baseline hazard. The explanatory variables act multiplicatively on the baseline hazard that is completely unspecified, and the regression parameter  $\beta$  in the model can be estimated by maximizing the partial likelihood function;

$$L(\beta) = \prod_j \frac{\exp(\beta' X_j)}{\sum_{i \in R(t_j)} \exp(\beta' X_{(i)})} \quad (5)$$

where  $R(t_j)$  denote the risk set of patients at time  $t_j$

### 2.4. Accelerated Failure Time Model (AFTM)

Accelerated Failure Time Model[19] assumes a linear relationship between logarithm of survival time and covariates, is given by

$$\text{Log} T_i = \mu + \beta' X + \sigma \omega \quad (6)$$

where  $\beta'$  is a vector of regression coefficients,  $\mu$  and  $\sigma$  are intercept and scale parameters respectively and the error term  $\omega$  is assumed to have some distribution(i.e extreme value, normal or logistic). This transformation leads to the Weibull, Lognormal or log logistic AFT models for  $T_i$ [20].

Now, the survivor function of  $T_i$  is given by

$$S_i(t/X) = S_0[t \exp(\exp(\beta'X))] \quad (7)$$

The AFT models are fitted by using method of maximum likelihood. The likelihood of  $n$  observed survival times  $t_1, t_2, \dots, t_n$  is given by

$$L(\beta, \mu, \sigma) = \prod_{i=1}^n \{f_i(t_i)\}^{\delta_i} \{S_i(t_i)\}^{1-\delta_i}$$

where  $f_i(t_i)$  and  $S_i(t_i)$  are the density and survival function for  $i$ th patients at time  $t_i$ , and  $\delta_i$  is the event indicator function, such that

$$\delta_i = \begin{cases} 1 & \text{if the event is observed} \\ 0 & \text{if it is censored} \end{cases}$$

Results obtained from AFT models can be summarized in the exponentiated form as time ratio (i.e.  $TR = \exp(\beta)$ ) unlike Cox model hazard ratio. Thus  $TR > 1$ , associated with prolonged survival time and  $TR < 1$ , associated with a decrease in survival time.

## 2.5. Model Comparison

In order to compare parametric AFTM and semi parametric Cox PH Model, we have used Akaike Information Criterion (AIC). The AIC which is a measure of goodness of fit of an estimated statistical model is given by

$$AIC = -2 * \text{Log-likelihood} + 2(p+k)$$

where  $p$  is the number of covariates in the model,  $k=1$  for exponential and  $k=2$  for weibull and lognormal models. Model with smaller AIC regarded as a better model.

## 3. Data Sources

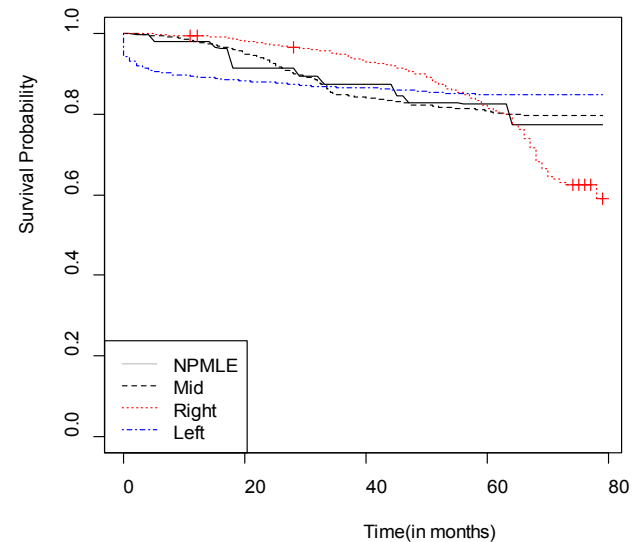
A retrospective follow-up study was conducted, involving 1259 HIV/AIDS patients who were undergoing Antiretroviral Therapy in the ART centre of Dr. Ram Manohar Lohia Hospital, New Delhi, India, during the period of April 2004 to November 2009, and were followed up through the ART routine register records till December 2010. Taking preliminary inclusion criteria as patients should be above 18 years of age and on the basis of availability of complete baseline information on CD4 cell count, date of visit, mode of transmission, weight and hemoglobin etc, 1259 patients were found eligible for the analysis. At the end of the study period, 198 patients were dead and remaining 1061 patients were known to be alive, since the exact date of death is not known, the event death is known to lie in an interval that is the last available CD4 count date and end of the study. Thus the observed event leads to interval censored survival data.

Table 1: presents the descriptive statistics of the study, out of 1259 patients 408 (32.4%) were female. The mean age at diagnosis was  $34.24(\pm 8.74)$  years and the most common mode of transmission was sexual (hetro+homo) 64.0%, and for 331(26.3%) patients mode of transmission were unknown. 858(68.1%) patients had CD4 cell count less than 200 cell/mm<sup>3</sup> at the time of enrollment. 198 (15.7%) patients died and remaining were alive at the end of the study.

**Table 1.** Descriptive Statistics

Variables	Category (Code)	N=1259	Percent
Sex	Male (0)	851	67.6
	Female (1)	408	32.4
Mode of Transmission (MOT)	Sexual (1)	806	64.0
	Blood+IDU (2)	122	9.7
	Unknown(3)	331	26.3
CD4+ cell	$\geq 200$ (0)	401	31.9
	$< 200$ (1)	858	68.1
Smoking	Current (1)	172	13.7
	Past (2)	294	23.3
	Never (3)	793	63.0
Status	Alive (0)	1061	84.3
	Death (1)	198	15.7
Hemoglobin		10.97 $\pm$ 1.87	Mean $\pm$ SD
Weight		50.10 $\pm$ 10.5	Mean $\pm$ SD
Age (in years)		34.24 $\pm$ 8.24	Mean $\pm$ SD
CD4+ count		192.06 $\pm$ 105.38	Mean $\pm$ SD

The Figure 1: shows NPMLE, K-M survival plot for left, right and mid-point imputation. Comparing these curves we can see that the K-M survival plot for mid-point imputation and NPMLE for intervals are very similar. Then assuming these data as right censored at mid-point of each interval, Cox Proportional Hazard Model and AFTM have been employed to study the effect of prognostic factors on the survival time.



**Figure 1.** Estimated survival function based on intervals and imputation

Table 2: presents the results of Cox PH model. The prognostic factors viz. sex, CD4 cell count, past smokers, baseline hemoglobin and baseline weight were found to be statistically significant ( $P < 0.000$ ). Patients who had CD4 cell count less than 200 mg/mm<sup>3</sup> ( $HR=3.31$ , 95% CI: 2.19, 4.98), sexually infected ( $HR=1.12$ , 95% CI: 0.75, 1.65) and current smokers ( $HR=2.14$ , 95% CI: 1.37, 3.34) had an increased hazard for death.

**Table 2.** Cox-PH Model for HIV/AIDS patients

Variables	$\beta$	HR	95% CI	P-value
Age	-0.012	0.987	(0.97, 1.02)	0.130
Sex	-0.780	0.458	(0.32, 0.64)	0.000***
CD4	1.197	3.311	(2.19, 4.98)	0.000***
MOT				
Sexual		1		
Blood+IDU	0.113	1.120	(0.75, 1.65)	0.570
Unknown	-0.185	0.830	(0.60, 1.14)	0.258
Smoking				
Current		1		
Past	0.760	2.140	(1.37, 3.34)	0.000***
Never	0.177	1.193	(0.75, 1.89)	0.451
HB	-0.201	0.817	(0.76, 0.87)	0.000***
Weight	-0.066	0.935	(0.92, 0.96)	0.000***
AIC			3628.829	

\*indicates significance at the 5% level, \*\* at 1% level and \*\*\* at 0.1% level

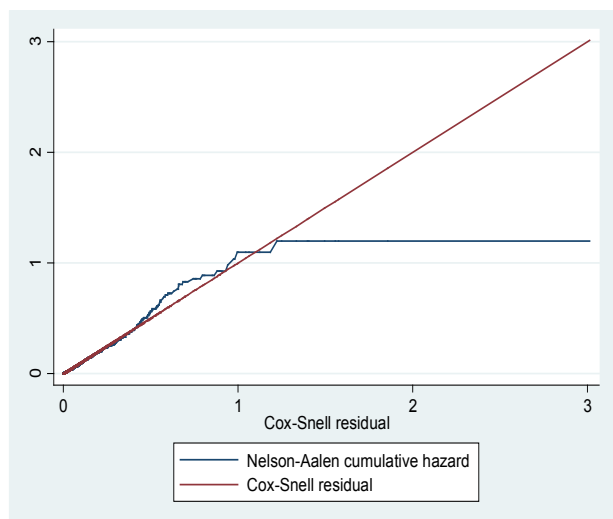


Figure 2.1. Cox PH Model

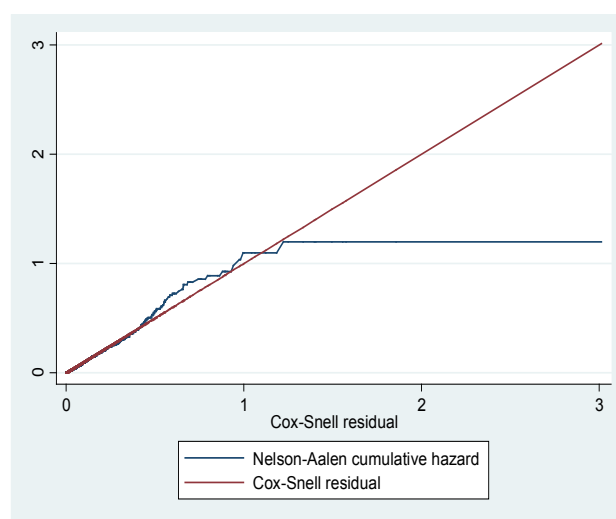


Figure 2.2. Exponential AFT Model

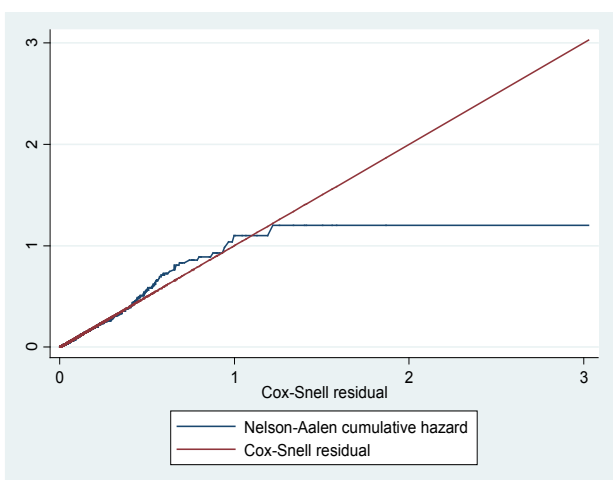


Figure 2.3. Weibull AFT Model

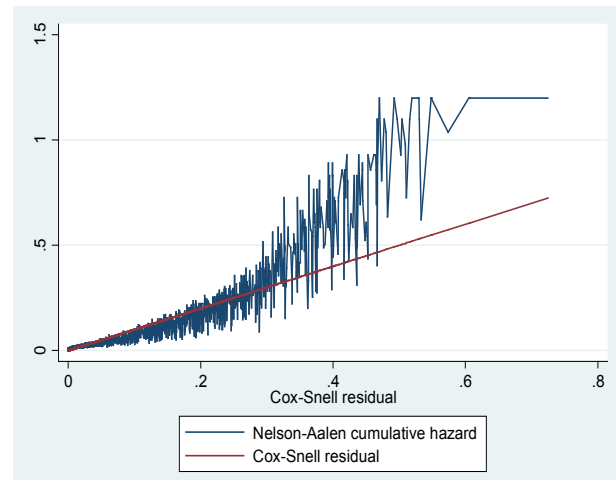


Figure 2.4. Lognormal AFT Model

**Figure 2.** Cumulative hazard plot of Cox-Snell residual for AFT models

**Table 3.** Parametric Accelerated Failure Time Model (AFTM) for HIV/AIDS patients

Parameters	Exponential			Weibull			Lognormal		
	$\beta$	TR	95% CI	$\beta$	TR	95% CI	$\beta$	TR	95% CI
Age	0.012	1.012	(0.99, 1.08)	0.008	1.008	(0.99, 1.03)	0.010	1.010	(0.98, 1.04)
Sex	0.734	2.083	(1.47, 2.94)**	0.534	1.707	(1.34, 2.16)**	0.507	1.661	(1.29, 2.13)**
CD4	-1.220	0.295	(0.19, 0.44)**	-0.820	0.440	(0.32, 0.59)**	-0.797	0.450	(0.34, 0.58)**
MOT									
Sexual		1			1			1	
Blood+IDU	-0.152	0.858	(0.58, 1.27)	-0.086	0.916	(0.70, 1.19)	-0.162	0.850	(0.62, 1.15)
Unknown	0.171	1.186	(0.86, 1.64)	0.239	1.270	(1.02, 1.58)**	0.333	1.395	(1.10, 1.77)**
Smoking									
Current		1			1			1	
Past	-0.769	0.463	(0.29, 0.72)**	-0.562	0.569	(0.41, 0.78)**	-0.559	0.571	(0.41, 0.79)**
Never	-0.138	0.870	(0.54, 1.38)	-0.132	0.875	(0.64, 1.19)	-0.095	0.909	(0.65, 1.25)
HB	0.199	1.220	(1.14, 1.30)**	0.144	1.155	(1.10, 1.21)**	0.145	1.156	(1.09, 1.21)**
Weight	0.062	1.066	(1.04, 1.09)**	0.046	1.047	(1.03, 1.06)**	0.042	1.043	(1.02, 1.06)**
Intercept	1.159			1.772			1.726		
$\lambda$				0.683			1.21		
$\gamma$	1			1.464					
AIC	<b>3359.18</b>			<b>3316.122</b>			<b>3322.259</b>		

\*indicates significance at the 5% level and \*\* at 1% level, Time Ratio (TR), TR<1, indicates a decrease in survival time and TR>1 indicates a prolonged survival time

The parametric AFTM results have been presented in table 3, besides the factors found significant in Cox model, unknown mode of transmission was also found to be a significant predictor in Weibull and lognormal AFT model. Patients who had CD4 cell count<200 cells/mm<sup>3</sup> have shorter survival time than patients with CD4 cell count more than 200 cells/mm<sup>3</sup> (as TR<1), female patients had longer survival than their male counterparts (TR>1). An increase in survival time was associated with per unit increase in hemoglobin. Old aged patients had shorter survival than young patients, and also an increase in weight (in kg) leads to increase in life expectancy.

According to Akaike Information Criterion (AIC), Weibull AFTM was found to be better (smallest AIC) one among the parametric and Cox model. Proportional assumptions were hold (confirmed by Schoenfeld residual plot, not shown here) and the Cox-Snell residual fitted well to all the AFT models shown figure 2.

## 4. Discussion

In this study we have tried to estimate the survival of HIV/AIDS patients under interval censoring mechanism, since the event of interest death lie in an interval i.e the last available CD4 count date and end of the study. Due to lack of well known statistical methodology and available statistical software, we have adopted imputation approach to handle interval censored data, so that the interval censored data converts to right censored data to which the standard technique can be applied. We have observed that survival function obtained by NPMLE (E-M) for interval censored data has a very close resemblance to the survival function obtained at the mid-point of the interval by Kaplan-Meier method. This is corroborated with the findings of previous studies[10, 17]. However our aim was to determine the effects of prognostic factors age, sex, MOT, baseline CD4

cell, baseline hemoglobin, baseline weight and smoking habits on the survival.

The prognostic factors viz. sex, CD4 cell count, past smokers, baseline hemoglobin and baseline weight are found to be statistically significant (P< 0.000) by both semi parametric Cox PH and parametric AFT Model. Most of the previous studies have suggested that the age is a significant prognostic factor[21, 22, 23]. As age increases the survival time of HIV/AIDS patients decreases. Old age is associated with high risk of disease progression but in our analysis age is not found to be a significant prognostic factor. Also females are observed to have better survival than their male counterpart. As reported previously female had higher life expectancies than male[21, 24, 25, 26]. Remafedi et al.[27], had proposed that sex of the patients does not have any significant effect on survival time.

Consistent with the published literature CD4 cell count is found to be an important prognostic marker of HIV/AIDS patients. Patients with CD4 cell <200 cells/mm<sup>3</sup> have 3.31 times more hazard to die than patient who had CD4 cell more than 200 cells/mm<sup>3</sup>. The mortality is inversely proportional to CD4 count, cumulative probability of AIDS and death increased substantially with decreasing CD4 cell count[23, 28, 29].

Another important result of our study is that patients with sexual (hetero or homo) mode of transmission had worst survival than patients with blood and intravenous drug user mode of transmission, which is departed from earlier result of[21]. where they found that intravenous drug user mode of infection had worst survival. However, Remafedi et al.[27] have shown that there were no significant differences between deceased and other subjects in relation to mode of transmission. Unknown mode of transmission was found to be a significant factor in weibull AFTM and Lognormal AFTM result.

Patients who are current smokers had association with more hazards of immune deterioration. We found that

baseline hemoglobin was a significant predictor of HIV/AIDS patients. To our knowledge hemoglobin has never shown to predict mortality of patients on ART in India, further studies needed to confirm our findings. Baseline hemoglobin level can be used as a simple and practical tool for initial risk assessment in the absence of CD4 cell count and viral load, as was identified in earlier studies by [30] in Tanzania and [31] in Europe. Patient's weight is positively associated with survival, this is corroborating to the findings of [22] that patients improved clinically with regard to weight and hemoglobin.

There are many situations where AFTM provides better description of data than Cox PH model [32, 33, 34]. Our aim of this study was to compare the performance of both semi-parametric Cox PH model and Parametric AFTM. Based on AIC, Weibull AFT model is found to be efficient (smallest AIC) among the parametric and Cox PH model.

There are some limitations of our study, firstly complete information on patients treatment, follow up data on clinical parameter were not available. Secondly, we have analyzed only one ART centre data, so generalization of our findings would need further conformity.

## REFERENCES

- [1] Williams B, Lima V and Gouws E.(2011) Modelling the Impact of Antiretroviral Therapy on the Epidemic of HIV. *Current HIV Research*. 9(6):367-382.
- [2] Stringer JSA, Zulu I, Levy J et al.,(2006). Rapid scale-up of antiretroviral therapy at primary care sites in Zambia: feasibility and early outcomes. *JAMA*. 296(7): 782-93.
- [3] Mahy M, Stover J, Stanecki K, Stoneburner R and Tassie JM.(2010). Estimating the impact of antiretroviral therapy: regional and global estimates of life years gained among adults, *Sex Transm Infect*. 86(Suppl 2):67-71.
- [4] National Aids Control Organization (NACO), Annual Report, 2011-12. Available at [http://www.nacoonline.org/upload/Publication/Annual%20Report/NACO\\_AR\\_Eng%202011-12.pdf](http://www.nacoonline.org/upload/Publication/Annual%20Report/NACO_AR_Eng%202011-12.pdf) (Accessed on 8th July 2012).
- [5] Turnbull B.W (1974). The Empirical Distribution Function with Arbitrary Grouped, Censored and Truncated Data, *Journal of Royal Statistical Society, series B*, 38,290-295.
- [6] De Gruttola V and Lagakos S. (1989). Analysis of doubly censored survival data, with application to AIDS. *Biometrics*, 45, 1-11.
- [7] Sun J. (1996) A non-parametric test for interval censored failure time data with application to AIDS studies. *Statist Med*, 15; 1387-1395.
- [8] Sun J. (2006). *The Statistical Analysis of Interval Censored Failure Time Data*. Springer, New York.
- [9] Kim M.Y and Xue X, (2002). The analysis of multivariate interval censored survival data, *Statistics in Medicine*, 21: 3715-3726.
- [10] Grover G and Banerjee T (2011). Estimation of Survival Times of HIV-1 Infected Children for doubly censored data, *Electron. J.App.Sta.Anal*, vol.4, issue 2,155-163.
- [11] Grover G and Shakeri N. (2007). Non parametric estimation of survival function of HIV+ patients with doubly censored data. *J. Communicable Dis*. 39(1): 7-12.
- [12] Gentleman R and Geyer C.J.(1994). Maximum likelihood for interval censored data: Consistency and Computation. *Biometrika*. 81, 618-623.
- [13] Wei L.J (1992) The Accelerated failure time model: a useful alternative to cox regression model in survival analysis. *Statistics in Medicine*, 11, 1871-1879.
- [14] Dempster A.P, Laird, N.M and Rubin D.B (1977). Maximum Likelihood from incomplete Data via the EM Algorithm, *Journal of the Royal Statistical Society, Series B*, 39, 1-38.
- [15] Hsu C.H, Taylor J.M.G and Murray S (2004). Multiple Imputation For Interval Censored Data With Auxiliary Variables, Working paper; Berkeley Electronic Press, paper 26.
- [16] Brookmeyer, R and Goedert J.J (1989). Censoring in an epidemic with an application to hemophilia-associated AIDS, *Biometrics*, 45, 325-335.
- [17] Law, C and Brookmeyer, R. (1992). Effects of midpoint imputation on the analysis of doubly censored data, *Statistics in Medicine*, 11: 1569-1578.
- [18] Cox D.R (1972) Regression Models and Life Tables, *Journal of Royal Statistical Society, series B*, 20,187-220.
- [19] Kalbfleish J.D and Prentice R.L. (1980). *The Statistical Analysis of Failure Time Data*: Wiley, New York.
- [20] Collett D (2003). *Modeling Survival Data in Medical Research*, Chapman and Hall, London.
- [21] Gadpayle AK, Kumar N, Duggal A et al., (2012) Survival trend and prognostic outcome of AIDS patients according to age, sex, stages and mode of transmission-A retrospective study at ART centre of a tertiary care hospital. *JIACM*, 13(4): 291-8.
- [22] Bachani D, Garg R, Rewari BB, et.al.(2010). Two year treatment outcomes of patients enrolled in india's national first-line antiretroviral therapy programme, *National Medical Journal of India*. 23(1):7-12.
- [23] Ghate M, Deshpande S, Tripathy S, et al.(2011). Mortality in HIV infected individuals in Pune, India. *Indian J Med Res*. 133:414-420.
- [24] Antiretroviral Therapy Cohort Collaboration.(2008). Life expectance of individuals on combination antiretroviral therapy in high income countries: a collaborative analysis of 14 cohort studies. *Lancet*, 372(9635): 293-9.
- [25] Farzadegan H, Hoover DR, Astemborski J, et al.(1998). Sex differences in HIV1 viral load and progression to AIDS. *Lancet*. 352:1510-1514.
- [26] Donnelly CA, Bartley LM, Ghani AC, et al.(2005). Gender differences in HIV1 RNA viral loads. *HIV Medicine*. 6(3):170-178.
- [27] Remafedi G, Lauer T.(1995). Survival trends in adolescents with human immunodeficiency virus infection. *Arch Pediatr Adolesc Med*. 149(10):1093-8.

- [28] Antiretroviral Therapy Cohort Collaboration.(2009). Timing of initiation of antiretroviral therapy in AIDS-free HIV-1 infected patients: a collaborative analysis of 18-HIV cohort studies. *Lancet* 6736(09) 60612-7.
- [29] May M, Sterne JA, Sabin C et al.,(2007). Prognosis of HIV-1 infected patients up to 5 years after initiation of HAART: collaborative analysis of prospective studies. *AIDS* 21; 1185-97.
- [30] Johannessen A, Naman E, Ngowi EN et al.,(2008). Predictors of mortality in HIV-infected patients starting antiretroviral therapy in a rural hospital in Tanzania, *BMC Infectious Disease*, 8: 52.
- [31] Mocroft A, Kirk O, Barton SE et al.,(1999). Anaemia is an independent predictive marker for clinical prognosis in HIV-infected patients from across Europe. EuroSIDA study group, *AIDS*, 13: 943-950.
- [32] Kay R and Kinnersley N (2002). On the use of accelerated failure time model as an alternative to the proportional hazard model in the treatment of time to event data: a case study in influenza, *Drug information Journal*, vol-36, pp; 571-579.
- [33] Orbe J, Ferreira E and Nunez-Anton V,(2002). Comparing proportional hazards and accelerated failure time models for survival analysis. *Statist Med*, 21; 3493-3510.
- [34] Nardi A and Schemper M (2003). Comparing Cox and Parametric models in clinical studies, *Statist. Med*, 22: 3597-3610.